NETWORK ANALYSIS OF COMPLIANCE ENFORCEMENTS

Brief introduction

Background of the Project

This is a data analytics field project of Fidelity AI Incubator (NLP department). I am working on this project since 06/2020. The project is anticipated to end in late August 2020.

Fidelity Investments delivers services to customers in a highly regulated environment; with regulatory agencies at the federal and state levels who establish rules, investigate potential violations, impose fines, and request information through inquiries. To ensure firm compliance, Fidelity Compliance Team analyzed fines imposed on other firms, update internal policies to ensure we comply with new or updated rules and respond to inquiry requests. Today this work is reactive, manual, and time consuming.

In this field project, Brandeis University team cooperated with Fidelity Compliance team to better understand connections between regulators, rules, firms, individuals, violations and fines imposed via data & business intelligence.

Goals of the Project

1) Build a visual static and an interactive semantic network with 6 nodes: regulators, rules, firms, individuals, fines imposed and violations based on the SEC filings about enforcement action in the past 7 years.
2) apply the relationships shown by the semantic network to identify patterns and trends of fines, based on which build predictive models.

Project Achievements

1. Used python to convert SEC filings about enforcement action (in xml format) to csv files. (1038 files，33370 enforcement records)
   [code] [sample file]
   (The raw xml files are confidential and belongs to Fidelity. I just show a sample data here, which contains the original structure of the xml file (do not change the nodes of the tree). The content in each node has been replaced or simplified.)
2. Built a relational database in MYSQL to store those enforcement files.
   [ER diagram]
3. Performed queries in MYSQL to get some basic patterns or regulations of the enforcement date; and extracted the enforcement data about fines or violations from the database for next step.
4. Applied python to give summaries in each enforcement record a series of tags, including the **name of the company/individual** which is imposed fines or violate a certain rule and the **amounts of fines or penalties** mentioned in each summary.
5. Build a visual static and an interactive semantic network with 6 nodes: regulators, rules, firms, individuals, fines imposed and violations based on the data.